Theory and Practice of Data Assimilation in Ocean Modeling

Robert N. Miller
College of Oceanic and Atmospheric Sciences
Oregon State University
Oceanography Admin. Bldg. 104
Corvallis, OR 97331-5503
Phone: (541) 737-4555 Fax: (541) 737-2064 Email: miller@coas.oregonstate.edu

Award Number: N00014-04-1-0262

LONG-TERM GOALS

The long-range goal of this project is to combine computational models with observational data to form the best picture of the ocean as an evolving system, and use that picture to understand the physical processes that govern the ocean's behavior. Oceanic observations are sparse and models are limited in accuracy, but taken together, one can form a quantitative description of the state of the ocean that is superior to any based on either models or data alone. Along with the goals of analysis and prediction, we seek reliable estimates of the errors in our results. We expect our results to have implications beyond data assimilation. In particular, we believe this research will lead to enhanced understanding of the implications of nonlinearity and randomness for predictability of the ocean and atmosphere.

In keeping with our goal of providing reliable error estimates for our data assimilation products, we seek to develop efficient methods for estimating useful statistical measures of errors in stochastic forecast models. Since the probability density functions (PDF's) of nonlinear stochastic models are not, in general, Gaussian, we must find methods for forecast evaluation based on information about the particular PDF generated by the model.

Since our goal is the development of practical analysis and forecast systems for the ocean, we want to solve remaining scientific problems involved in transition from data assimilation experiments tuned to specific models and data sets to operational analysis and prediction on a research basis. This will involve rigorous quantification of the information content of each data set, as well as quality control, a problem with which the ocean modeling community has limited experience.

OBJECTIVES

The principal objective of this project is the development, implementation and evaluation of practical data assimilation methods for regional to basin scale ocean models. By "data assimilation" we mean the construction of a composite estimate of the state of the ocean based on a combination of observed data with computational model output. Since data assimilation methods which give the most and best information are highly resource intensive, and often not practical for use with detailed models, we are particularly interested in the price paid in terms of accuracy and confidence for using economical but suboptimal data assimilation methods.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **30 SEP 2006**	2. REPORT TYPE	3. DATES COVERED **00-00-2006 to 00-00-2006**
4. TITLE AND SUBTITLE **Theory and Practice of Data Assimilation in Ocean Modeling**		5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Oregon State University,College of Oceanic and Atmospheric Sciences,Corvallis,OR,97331**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release; distribution unlimited**		
13. SUPPLEMENTARY NOTES		
14. ABSTRACT		
15. SUBJECT TERMS		

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**7**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

Direct calculation of full PDF's is not feasible for practical models of the ocean or atmosphere, but useful approximations to the PDF can be calculated from Monte-Carlo experiments, by virtue of the fact that the number of truly independent degrees of freedom in practical models is very much smaller than the dimension of the state vector. This intuition is the motivation for the ensemble methods that have become popular in recent years.

Our experience with Monte-Carlo methods in simplified systems has led us to investigate the details of methods for ensemble generation that have been presented in the community. The motivation for these specialized methods for generating ensembles is precisely the specification of the PDF of a complex model whose behavior is believed to be captured by a relatively small number of independent degrees of freedom. By detailed study of the behavior of ensembles in increasingly complex models, we hope to gain the insights necessary to generate the most efficient ensembles, which should, in turn, lead to the error estimates necessary for data assimilation systems and prior estimates of forecast accuracy.

Optimized methods require accurate knowledge of the statistics of the errors in the model and the data. It is therefore an objective to understand in detail the sensitivity of the data assimilation scheme to the details of the defining error estimates.

APPROACH

The basic assumptions underlying data assimilation methods in use or proposed are known to be false to some degree. We plan to study the consequences of these assumptions by constructing a hierarchy of schemes with decreasing reliance on ad hoc assumptions. It is our guiding philosophy that the best way to learn how to design and implement the most economical methods that meet our needs is to begin by implementing methods which are as close to optimal as possible. From that point, we can quantify the loss of accuracy and the saving of resources associated with each simplification of the model or the data assimilation scheme.

Work is proceeding toward a theoretical basis for the next generation of data assimilation methods in which randomness and nonlinearity must be taken into account. To this end, we are applying tools from stochastic differential equations and from dynamical systems theory. Since our model systems are characterized by high dimensional state spaces, Monte Carlo methods must be used to study the behavior of the stochastic systems.

The theory of nonlinear filtering provides a framework in which problems of data assimilation with nonlinear models and non-Gaussian noise sources can be treated (see, e.g., Miller et al., 1999). In the case of linear models and Gaussian noise sources, this theory reduces to the familiar Kalman filter. In the formal theory of nonlinear filtering, the final result is not a single model state vector or trajectory in state space, but a PDF defined as a scalar function of the state variables and time. From this PDF, the mean, median, mode, or other statistic can be computed for use as the working estimate of the state of the system, along with the desired confidence intervals. The assignment of confidence limits corresponds in the case of a group of particles in physical space to drawing contours in the spatial domain which can be expected to define a region which contains, say, 90% of the particles.

The problem is that for even schematic models of the ocean or atmosphere, an unrealistically large number of particle trajectories in phase space must be calculated in order to represent the PDF faithfully. Useful ensemble analysis therefore requires judicious choice of ensemble members. We

have concentrated our recent efforts on evaluation of ensemble methods, which we see as facilitating the generation of the forecast error estimates necessary for data assimilation. These forecast error estimates are of interest in and of themselves, since they have the potential of providing a priori estimates of the reliability of a given forecast.

We are now in the process of investigating the qualitative behavior of more complex and relevant systems. Results from the theory of dynamical systems lead to methods for explicit construction of the low dimensional spaces in which meaningful probabilistic calculations can be performed on complex systems. At this moment, our work is focused on models of the Kuroshio. The simplest of these is a regional quasigeostrophic model which reproduces the observed bimodality. It operates on a state space with several thousand dimensions. This is two orders of magnitude greater than that of earlier schematic models, and, for this reason alone, presents significant technical challenges.

We now have a basis of comparison with more complex models, up to and including eddy resolving primitive equation models of the north Pacific. We are now in the process of applying our methods from dynamical systems and stochastic calculus to a suite of models, in order to understand propagation of errors and the evolution of the PDF arising from random initial and boundary conditions in a state space of workable dimension. This should allow us to construct reliable data assimilation systems for use with simulated and real data from the Kuroshio.

Many different models, based on fundamentally different physical assumptions, exhibit the observed bimodality of the Kuroshio in some form. We are now in the process of comparing our model to different models and to observed data in order to determine a basis for distinction between the physical mechanisms in the different models.

Professor James G. Richman of the College of Oceanic and Atmospheric Sciences at Oregon State is working with us on the dynamical analysis of the Kuroshio. Professor Richman is working on analysis of time series, and on comparison of ocean general circulation model results to data and to other models. Technical support for this project is provided by Ms. Laura Ehret. Guillaume Vernières successfully applied inverse theory to a regional baroclinic quasigeostrophic model of the Kuroshio. He completed his doctoral work in the winter of 2006, and is now a postdoctoral fellow with Chris Jones in the Department of Mathematics at UNC Chapel Hill and at the Statistical and Applied Mathematical Sciences Institute in Research Triangle Park, NC.

WORK COMPLETED

We have implemented a generalized inverse model with a baroclinic quasigeostrophic model with boundary-fitted coordinates and used it to provide initial conditions for a year-long forecast by assimilating satellite altimetric data; see Figure 1.

We have developed a method for projection of the results of a 1/10o primitive equation simulation of the north Pacific (provided by J. McClean and M. Maltrud) on the state space of our quasigeostrophic model. To do this we have implemented a method for decomposing the velocity field of the primitive equation model into irrotational and nondivergent components and calculating a stream function for the nondivergent part. The divergent part can be viewed as a diagnostic for the validity of the QG model. This is an example of a projection method (e.g., Brown et al., 2001); see Figure 2.

RESULTS

Assimilation of altimetric data for three weeks can provide initial conditions to a quasigeostrophic forecast model, which will predict the formation the Kuroshio eddy off the coast of Japan. The eddy was not present in the first guess field. The forecast is skillful for well over a year. The same system will not predict the decay of the eddy.

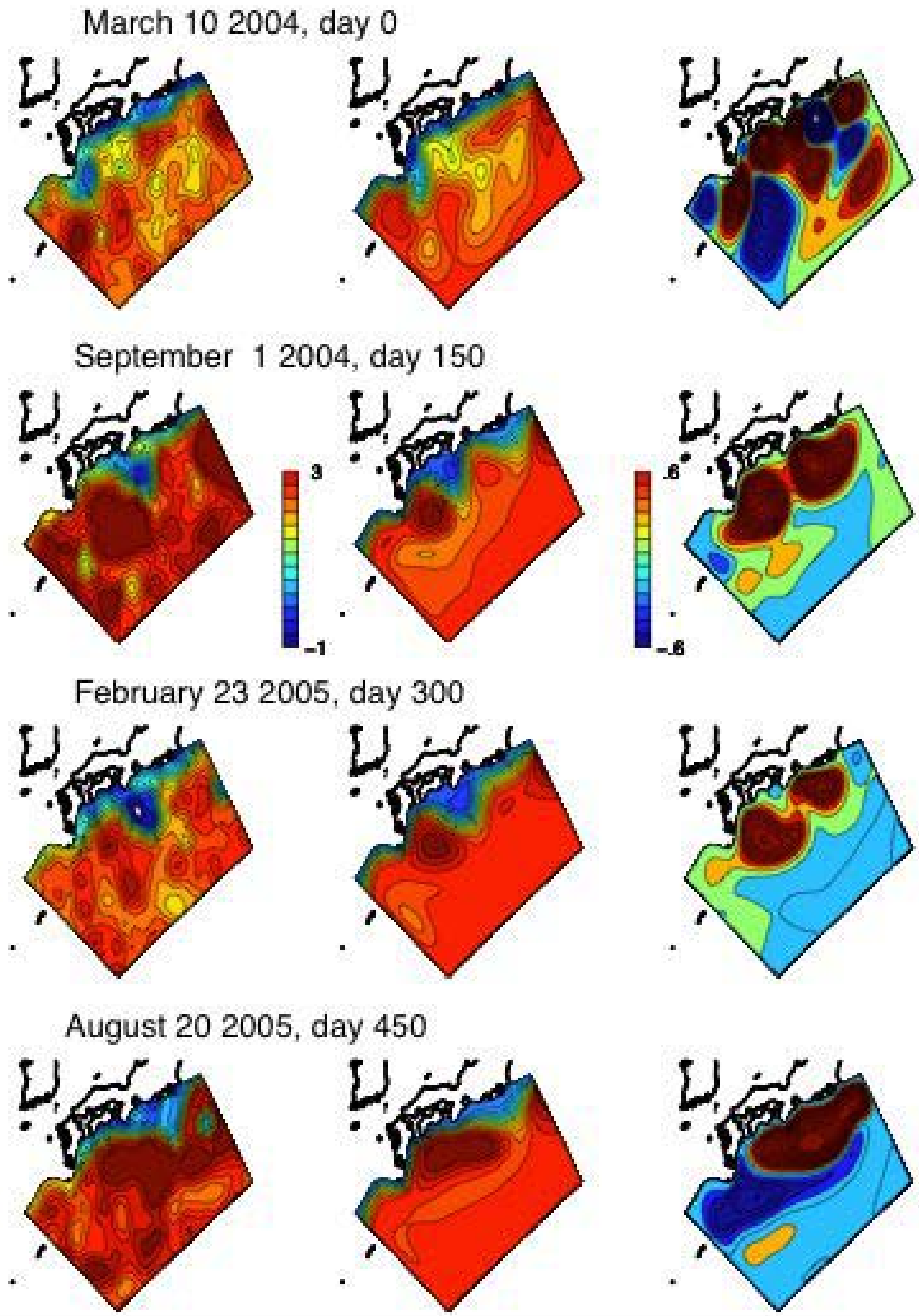

Figure 1. Forecast of Kuroshio meander formation. Left panels: Satellite sea surface height, scaled as upper level quasigeostrophic streamfunction. Middle panels: Model upper layer streamfunction. Right panels: Model lower level quasigeostrophic streamfunction. Top panels, middle and right column: Model initial condition, derived from assimilation of satellite altimetric data for an 18 day period ending March 10, 2004.

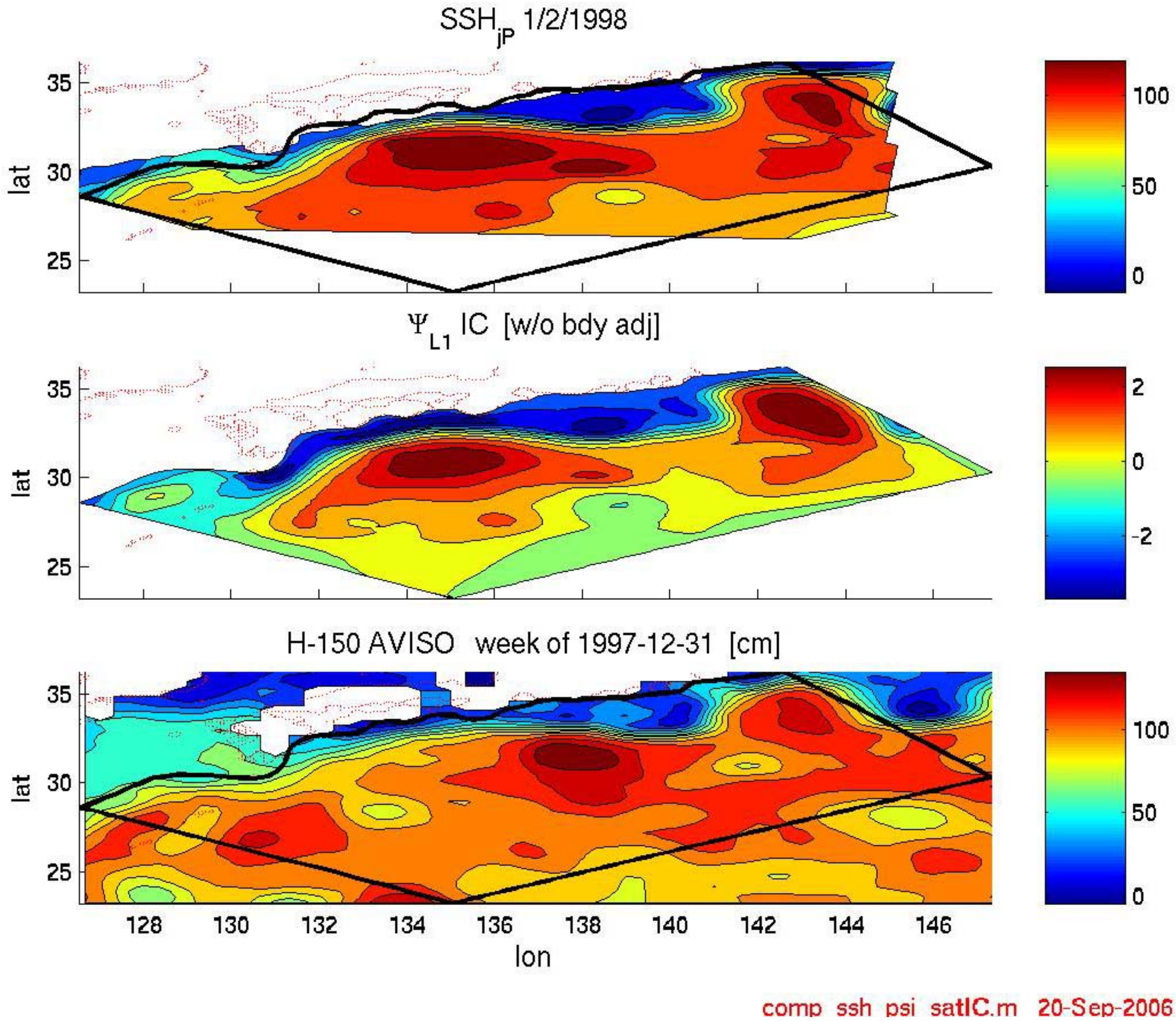

Figure 2. Comparison among model sea surface height, model streamfunction derived by the projection method, and observed sea surface height for the Pacific ocean off the coast of Japan. Top: Sea surface height from output of the 1/10o primitive equation model of J. McClean and M. Maltrud for January 2, 1998. Center: Streamfunction calculated from the nondivergent component of the 1000m depth-averaged velocity field from the 1/10o primitive equation model. Bottom: Sea surface height analysis from the AVISO satellite altimetry product for the week of December 31,1997.

We have evaluated the output of the 1/10 degree primitive equation model, and found that projection of the 1000m depth-averaged velocity field onto its nondivergent component results in a streamfunction that is comparable to the sea surface height fields produced by the model and observed by satellite altimetry. The ageostrophic velocities are small in the region of interest, despite the relatively steep topography.

IMPACT/APPLICATIONS

Major weather centers, including the US National Center for Environmental Prediction (NCEP) and the European Center for Medium-Range Weather Forecasting (ECMWF) now use ensemble methods for operational forecast validation; see Molteni et al. (1996), Toth and Kalnay (1993). Our work on Monte-Carlo methods should provide enhanced capability for validation of forecasts of the ocean and atmosphere, in addition to application to data assimilation. Our work on breeding modes and planned work on other schemes for ensemble generation should provide significant guidance in optimizing methods for generation of ensembles. Our work on dynamical analysis of models of the Kuroshio should lead to practical methods for identification of low-dimensional spaces in which efficient ensemble methods could be implemented.

We expect our the results of our inverse model of the Kuroshio to shed light on the importance of nonlinearity in ocean models; further, we expect that our work with comparisons among models and data for the Kuroshio will lead to greater insight into the intrinsic variability of basin-scale ocean circulation.

RELATED PROJECTS

"Assimilation of Coastal Radar Surface Current Measurements in Shelf Circulation Models." Work is in progress on the investigation of data assimilation systems for use with surface velocity data from coastal radar. This project is in collaboration with Professors Alexandre Kurapov, John Allen and Gary Egbert.

REFERENCES

Brown, D. L., R. Cortez and M. L. Minion, 2001: Accurate projection methods for the incompressible Navier-Stokes equations. *J. Comput. Phys.* **168**, 464-499.

Miller, R. N., E. F. Carter and S. T. Blue, 1999: Data assimilation into nonlinear stochastic models. *Tellus*, **51A**, 167-194**.**

Molteni, F., R. Buizza, T. N. Palmer and T. Petroliagis, 1996: The ECMWF ensemble prediction system: Methodology and validation. *Q. J. R. Meteorol. Soc.*, **122**, 73-119.

Richman, J. G., R. N. Miller and Y. H. Spitz, 2005: Error estimates for assimilation of satellite sea surface temperature data in ocean climate models. *Geophys. Res. Lett.* **32**, L18608, doi:10.1029/2005GL023591

Toth, Z. and E. Kalnay, 1993: Ensemble forecasting at NMC: the generation of perturbations. *Bull. Am. Meteorol. Soc.*, **74**, 2317-2330.

PUBLICATIONS

Miller, R. N. (2006) Topics in Data Assimilation : Stochastic Processes. *Physica D*, in press. [in press, refereed]

Miller, R.N. (2006) *Numerical Modeling of Ocean Circulation*. Cambridge University Press. [book, in press].